ZHAOKUN WANG

AI Engineer | LLM Agents & Enterprise Software

wang@cl.uni-heidelberg.de | +49 152 09723758 | Heidelberg, Germany
www.github.com/BufferHund | www.iamzhaokun.com
www.linkedin.com/in/zhaokun-wang-8802452a9

EDUCATION

M.A. Computational Linguistics, AI Track

Heidelberg University

Sep. 2024 – Oct. 2026 (Expected) | Heidelberg, Germany

- GPA: **1.5**. Focus on agentic AI, multimodal language models, VLM-based grounding, knowledge-enhanced systems, and human–AI interaction.
- Selected coursework: Multimodal Language Models, Computer Vision, Knowledge Graphs, Temporal Reasoning, and AI-Assisted Software Engineering.

B.Eng. Computer Science

Central South University of Forestry and Technology

Sep. 2019 – Jun. 2023 | Hunan, China

- GPA: **1.8**. Established a strong foundation in algorithms, data structures, databases, and software engineering.

PROFESSIONAL EXPERIENCE

AI Research Intern / Master's Thesis Student

Volkswagen AG & Group Innovation

Jan. 2026 – Present | Wolfsburg, Germany

- Work on **computer vision and vision-language models** for multimodal AI applications, with a focus on visual understanding, grounding, and interaction in complex environments.
- Conduct Master's thesis research on **AI agent safety domain**, focusing on the evaluation of agent behavior, interaction, and reliability.
- Contribute to **Agent Harness development** for multimodal AI systems.

Software Engineer

Shenzhen S.C New Energy Technology Corporation

Jul. 2023 – Jun. 2024 | Shenzhen, China

- Implemented and maintained MES functions using TypeScript and C#, supporting manufacturing workflow management and production data handling.
- Translated shop-floor and business requirements into reliable software features for reporting, workflow configuration, and operational support.
- Collaborated with product, QA, and operations teams in Agile delivery, including testing, issue tracking, and deployment support.

PROFILE

Master's candidate in Computational Linguistics with a Computer Science background and experience in **LLM agents, multimodal AI, and software engineering**. Builds reliable AI applications integrating models, APIs, tools, and enterprise workflows.

PROJECTS

Resource-Efficient Qwen Distillation

- Built a teacher–student distillation pipeline for **Qwen2.5-0.5B** using LoRA and 8-bit quantization.
- Retained **75% of teacher accuracy with 6× lower compute**.

Hierarchical Language Model with Cache

- Re-implemented a hierarchical character-level language model in **PyTorch** with vectorized computation and continuous cache management.

For more project details, please visit: https://www.iamzhaokun.com/research page.

TECHNICAL SKILLS

AI Agents & LLM Systems
LLMs, RAG, MCP, Tool Calling, Prompt Engineering, Context and Memory Management

Machine Learning
PyTorch, VLMs, LoRA, Quantization, Knowledge Distillation, Model Evaluation

Programming & Backend
Python, Java, TypeScript, C#, SQL, FastAPI, REST APIs, .NET

Cloud & Engineering
Azure, Docker, Linux, Git, CI/CD, VS Code, Cursor

LANGUAGES

English (Fluent)
German (Basic)
Chinese (Native)

AVAILABILITY

Available from October 2026